

REMARKS

Claims 1, 4-5, 7-11, 14-15, 17-23, and 29-32 are pending in the present application. Claims 1, 4-5, 7-11, 14-15, 17-23, and 29-32 were rejected under 35 U.S.C. §103(a).

Section 103 Rejections

The Examiner has rejected claims 1, 4-5, 7-11, 14-15, and 17-23 under 35 U.S.C. §103(a) as being obvious over U.S. Patent No. 6,081,780 (Lumelsky) in view of Applicant's Admitted Prior Art (AAPA).

The Examiner has rejected claims 29-32 under 35 U.S.C. §103(a) as being obvious over Lumelsky and AAPA, and further in view of Saon, et al., "Maximum Likelihood Discriminant Feature Spaces", IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, June 1000, pgs. 1129-1132

The Examiner specifically cited Lumelsky, col. 10, line 49, to col. 12, line 25, as disclosing a method for text-to-speech (TTS) synthesis that includes *providing a text string comprising a plurality of words and phonemes and corresponding spoken audio signal wherein a user specifies a pronunciation of the text string*, as essentially recited in claims 1, 11, and 21.

Applicant respectfully disagrees with this interpretation of Lumelsky.

Applicant has described Lumelsky in the responses filed on May 30, 2008 and December 19, 2008. Lumelsky is directed to enabling content providers with authoring tools to provide a highly compressed voice content. The section cited by the Examiner discloses a singlecast interactive radio system that offers a human-authored TTS system for improved quality of text-to-speech conversion. A decompression engine synthesizes a voice from a CES file, using one or more recorded allophone dictionaries which may be individually selected by the user. Allophones are variants of phonemes based on surrounding speech sounds. The allophones recorded in the dictionaries respectively define the preferred narrator voices, one of which may be chosen by the user. The user may preselect, via a

voice-command, the type of “voice” for narrating the requested decompressed information and, depending on the selection, the appropriate allophone dictionary is used to speech synthesize the information. By issuing corresponding commands, the user not only can choose among several of the dictionaries to be used interchangeably, but also can control the playback rate, level, repeat, fast forward, skip to the next file, and any other similar playback related functions, as will be further explained. Lumelskys’ authoring system can correct the intonation and add emotional content to the audio delivery. The authoring system embodies a speech processing system which compares audio produced by an operator (narrator), who reads the text aloud, with the speech synthesized artificially from the same text. The comparison results are used to improve a phonetic representation of the text. Subsequently, prosodic information is sent along with the phonetic representation of the text data to the customer terminal.

Thus, Lumelsky’s speech synthesis is based on an allophone context tables converter and one or more dictionaries, whereas the speech synthesis recited in Applicant’s claims 1, 11, and 21 is based on the spoken audio signal. Lumelsky teaches using a spoken audio signal to modify the pronunciation derived from the allophone dictionary, based on a prosody analysis (see Fig. 2B), but does not teach or suggest using a text string and the spoken audio signal to output a duration contour essentially as claimed. The speech synthesis recited in Applicant’s claims 1, 11 and 21 provides the speaker with the ability to override the output of Lumelsky’s allophone context tables converter, which is not taught or suggested by Lumelsky. In addition, Lumelsky extracts duration information from audio alone (see block 122 in Fig. 2A). Furthermore, the AAPA cited by the Examiner is directed to the Viterbi algorithm in general and not to “Viterbi algorithm . . . in the art”, as suggested in the rejection, and thus does not rectify this deficiency of Lumelsky. Indeed, the rejection fails to cite a reference disclosing alignment of text and speech. Thus, Applicant urges that the combination of Lumelsky and AAPA does not teach or suggest all limitations of independent claims 1, 11, and 21, and therefore that a *prima facie* case of obviousness of those claims over Lumelsky and AAPA cannot be maintained. Reconsideration and withdrawal of these rejections are respectfully requested.

Claims 4-5, 7-10, 14-15, 17-20, 22-23 all depend from either claims 1, 11, or 21, respectively, and are thus patentable for at least the same reasons as claims 1, 11, and 21. Reconsideration and withdrawal of these rejections are respectfully requested.

The Examiner cited Saon as disclosing the subject matter now recited in new claims 29-32. However, Saon is not directed to speech to text synthesis, and thus does not remedy the deficiencies of Lumelsky, discussed above. Reconsideration and withdrawal of these rejections are respectfully requested.

CONCLUSION

Applicant urges that claims 1, 4-5, 7-11, 14-15, 17-23 and 29-32 are in condition for allowance for at least the reasons stated. Early and favorable action on this case is respectfully requested.

Respectfully submitted,

By: David L. Heath Date: 5/26/09
David L. Heath
Reg. No. 46,763

Mailing Address:

**F. Chau & Associates, LLP
130 Woodbury Road
Woodbury NY 11797
(516) 692-8888
(516) 692-8889 (FAX)**